# Crowdsourcing for the Swedish Sign Language Dictionary

**Nikolaus Riemer Kankkonen, Thomas Björkstrand, Johanna Mesch, Carl Börstell**
Department of Linguistics, Stockholm University
S-106 91 Stockholm, Sweden
{nikolaus.kankkonen, bjorkstrand, johanna.mesch, calle}@ling.su.se

**Abstract**

In this paper, we describe how we are actively using the Swedish Sign Language (SSL) community in collecting and documenting signs and lexical variation for our language resources, particularly the online Swedish Sign Language Dictionary (SSLD). Apart from using the SSL Corpus as a source of input for new signs and lexical variation in the SSLD, we also involve the community in two ways: first, we interact with SSL signers directly at various venues, collecting signs and judgments about signs; second, we discuss sign usage, lexical variation, and sign formation with SSL signers on social media, particularly through a Facebook group in which we both actively engage in and monitor discussions about SSL. Through these channels, we are able to get direct feedback on our language documentation work and improve on what has become the main lexicographic resource for SSL. We describe the process of simultaneously using corpus data, judgment and elicitation data, and crowdsourcing and discussion groups for enhancing the SSLD, and give examples of findings pertaining to lexical variation resulting from this work.

**Keywords:** Swedisch Sign Language, lexicography, variation, sociolinguistics

## 1. Introduction

The Swedish Sign Language Dictionary (SSLD) (Svenskt teckenspråkslexikon, 2018) has been the main lexical database for Swedish Sign Language (SSL) since 2008. Initially constructed as an online video representation of an earlier printed dictionary (Hedberg et al., 1998), it soon grew to a limited resource following criteria of the printed dictionary.

An important aspect of any language resource is that it is representative of the language or register it covers. In the case of resources also serving as language documentation, this is perhaps even more important. There are currently two main language resources of SSL publicly available, a dictionary and a corpus (Mesch et al., 2012; Mesch and Wallin, 2012), both of which also serve as a form of language documentation for SSL. The first is the Swedish Sign Language Dictionary (SSLD), which is an online video dictionary, currently containing 17,310 entries, and in some cases sentence examples (available for around 22% of the entries, in total 3,944 sentence examples). Each dictionary entry is represented by a video of the sign (or phrase), a Swedish translation, phonological information, and internal cross-links to phonologically or semantically equivalent signs – i.e. homophones and synonyms (Svenskt teckenspråkslexikon, 2018). The second resource is the Swedish Sign Language Corpus (SSLC), which consists of 24 hours of video data (conversations, narratives, and presentations) from 42 different signers collected in the years 2009–2011 (Mesch et al., 2012; Mesch, 2018).

Although both the SSLD and the SSLC are designed and maintained by the Sign Language Section at the Department of Linguistics, Stockholm University, they have initially been set up independently from each other. This has led to some discrepancies between the functions of the two resources. However, the two projects have started to converge in the last year, and now they are already supporting each other. For example, sign tokens in the SSLC are annotated with the SSLD ID number (where matching is possible), signs are added to the SSLD as they appear in the SSLC, and the SSLD online interface sorts synonym search hits according to SSLC frequencies calculated through the SSLC lexical frequency interface (Börstell and Östling, 2016). However, seeing as the SSLC currently contains only around 90,000 sign tokens, and there was no explicit lexical variation elicitation task during the collection of the corpus data (Stamp et al., 2014), many sign synonyms or form variations cannot yet be investigated solely with the use of corpus data.[1] A set of 90,000 sign tokens is fairly large for being a sign language corpus, but small when compared to spoken language corpora used to investigate variation systematically. Instead, we make use of crowdsourcing to overcome some of these obstacles. In this paper, we present our methods for involving the language community to collect new signs and to gather data on familiarity with and variation within sign synonyms – that is, crowdsourcing to improve our language resources, specifically the SSLD. Thus, we include the community in enhancing the functionality of the resources later used by the community itself.

## 2. What Is the Sign for X?

The SSLD is an important resource for many different groups within the SSL community. It is used by SSL students, interpreters, and also Deaf signers as the go-to reference for looking up signs and sign variants in their language. Two common questions that are asked are: a) *what is the sign for X?*; and b) *which of the sign variants for X should I use?* Regarding the first question, it is an issue for language documentation: we need to find, document, and distribute knowledge about the signs that are used by the community. As for the second question, there is sometimes a general "hunch" as to which variant from a set of signs with more or less similar form is used by which sociolectal or dialectal group (e.g., *"This is an older sign"*, or, *"This*

[1] Approximately half of the collected 24 hours of data have been annotated, thus far.

*sign is used mostly in Stockholm"*), but we aim to provide a research-based and data-driven dictionary, which means we need more data to support such claims. In some cases, we may use the SSLC data and its frequency tool (Börstell and Östling, 2018) to compare the relative frequencies of two competing sign variants. An example of this is shown in Figures 1–2 with the sign TIO ('ten') in two different forms, a one-handed (Figure 3) and an older two-handed form (Figure 4).[2]

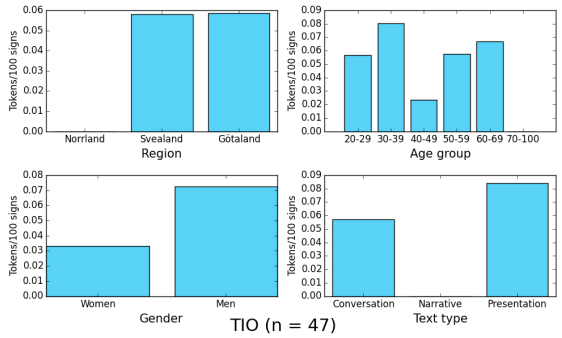


Figure 1: The relative frequencies of the one-handed sign TIO ('ten') – SSLD ID 4475

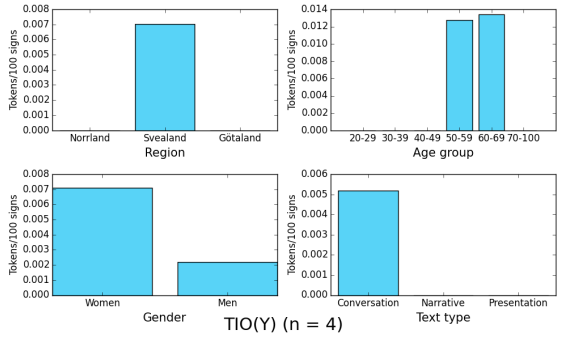


Figure 2: The relative frequencies of the two-handed sign TIO(Y) ('ten') – SSLD ID 11951

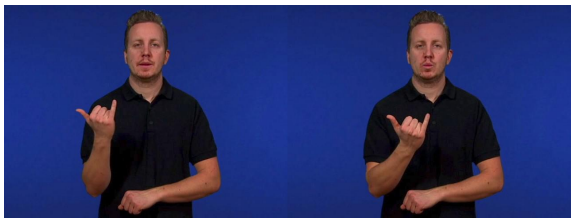

Figure 3: The one-handed sign TIO (SSLD ID 4475) ('ten')

As the figures show, the fact that TIO(Y) is an older sign than TIO is visible in their respective distributions: TIO

[2]The ID numbers in the figure captions are linked to the sign entry in the SSLD.

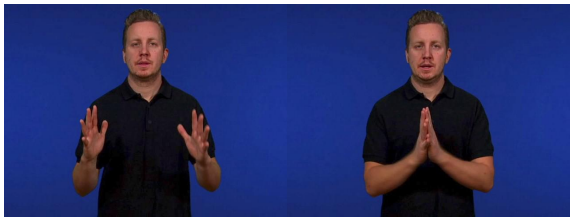

Figure 4: The two-handed sign TIO(Y) (SSLD ID 11951) ('ten')

is more common overall, and TIO(Y) is restricted to older signers. However, already here we are dealing with quite few data points (only four tokens for TIO(Y)), which is where our crowdsourcing comes into play.

We have two main crowdsourcing strategies: the first is a Facebook group entitled *Teckenspråkslexikon* ('sign language dictionary') administrated by our dictionary team; the second is our annual participation in *Dövas Dag* ('the Deaf Day'), which is a national convention for Deaf organizations and Deaf community businesses and activities.

## 3. The Facebook Group

Our Facebook group *Teckenspråkslexikon* was started in October 2014, after that year's Deaf Day, by the dictionary team. It currently has 2,642 members – see Figure 5. Figure 6 shows the monthly increase of members in the group from its start until now (October 2014–January 2018). It is visible here that the number of members greatly increases around two points during the year: September–October and January–February. We expect that this is due to two specific events. First, the Deaf Day is organized in September each year, when the SSLD team has informed about their work there. Second, our sign language courses at the Department of Linguistics, Stockholm University, start in September and January. The Facebook group has a diverse set of members, consisting of Deaf, hard-of-hearing, and hearing individuals, who are involved in the community in different ways (Deaf, Codas, interpreters, and friends and relatives of Deaf people, etc.).

The group is based around members asking for and discussing different signs, mainly under the premise that the meaning is not found in the online dictionary. Members then interact in different ways, discussing which signs are to be used, or the difference between sign variants. This interaction takes place both in written Swedish and in SSL by members uploading their own video comments, thanks to the video comment function in Facebook. In the last two years (i.e., January 2016 to January 2018), there have been 593 posts and a total of 5,817 interactions (original posts and comments combined) – see Figure 7. The peak phases of the period show that the posts and comments appear most in synchrony with peaks in member expansion – i.e., September–October and January–February (cf. Figure 6).

Table 1 shows the distribution of post types in the Facebook group from January 2016 to January 2018. As the table shows, 84% of the all posts consist of questions about

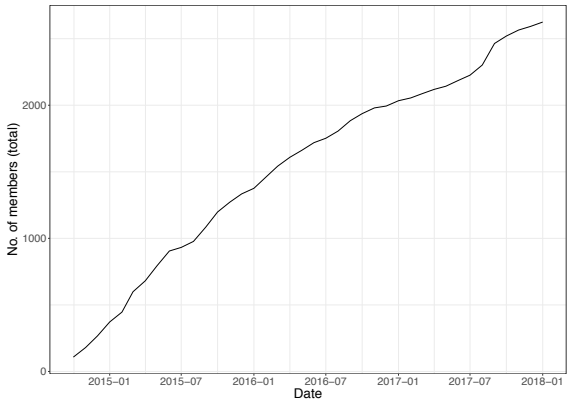


Figure 5: The number of total members in the Facebook group between October 2014 and January 2018

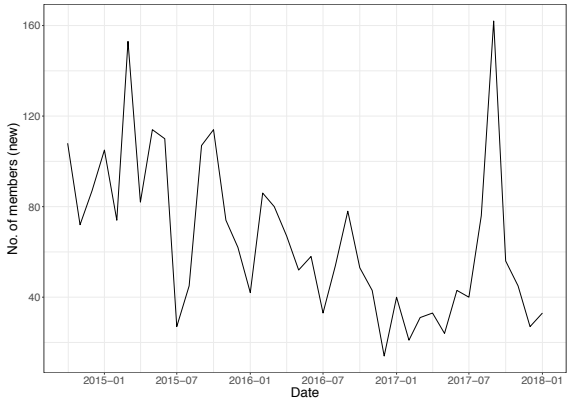


Figure 6: The number of new members joining the Facebook group per month between October 2014 and January 2018

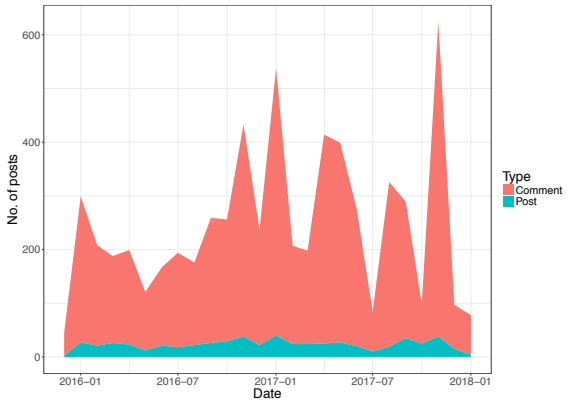


Figure 7: Amount of posts and comments of the Facebook group members between January 2016 and January 2018

signs. This includes questions about signs, such as *how do you sign X?* (53%), or name signs, such as *is there a sign for person Y?* (20%), but also general questions about the usage (10%) and etymology (<0.2%) of specific signs. 12% of the posts are about information, of which half are people giving information about something related to SSL or the SSLD (e.g., new features in the SSLD, or language resources), and the other half are requests for information about some topic (e.g., sign language courses or tools). Finally, 4% of the posts are not directly related to the SSLD or SSL, and are thus categorized as *Other*.

| **Post type** | **Number** | **%** |
|---|---|---|
| Question | 498 | 84% |
| – *sign* | *315* | *53%* |
| – *name sign* | *121* | *20%* |
| – *usage* | *61* | *10%* |
| – *etymology* | *1* | *<0.2%* |
| Information | 72 | 12% |
| – *giving* | *36* | *6%* |
| – *searching* | *36* | *6%* |
| Other | 23 | 4% |
| | 593 | 100% |

Table 1: Number of posts by type in the Facebook group (December 2015–January 2018

Of the posts concerning questions about signs, approximately half of the posts concern signs already found in the SSLD, whereas the other half do not. However, there are some signs requested that were already in pre-published stage of the SSLD, and many were added to the SSLD after the request was made, some of which were based on suggestions in the comments. Through the comment section, it is possible to follow the discussions of the community members and their reactions and ideas about signs in the SSLD or signs suggested in video comments. However, there are only few discussions about whether a sign entry in the SSLD is correct or not, and in those cases, approximately 90% of the commentators think that the SSLD sign entry is good.

Other discussions about signs may be about a meaning or the lexical category of a sign, for example TO-PROGRAM (verb) vs. PROGRAMMER (a person who writes computer software). Especially when signs for new concepts (e.g., technology) are demonstrated in video comments, we can, based on the reaction and informal ratings (e.g., reaction buttons), choose to include the sign in the dictionary, as part of language documentation and a way of enhancing the language resource for the public. Thus, the Facebook group helps us answer both questions (partially) – i.e., both *what* new signs there are, and *who* uses them. Many new sign entries have been added to the SSLD as a direct consequence of their being demonstrated in the Facebook group.

## 4. Direct Contact with the SSL Community

During our participation at the last Deaf Day in September 2017, we had devised a questionnaire in Google Forms with the intention of collecting variation data from community members. The questionnaire started with a set of background questions (i.e., signer metadata) followed by

25 questions of the type *"How do you sign X?"*, with all the documented sign variants available as animated `.gif` files.[3] Deaf attendees at the convention would participate in our study by responding to the questionnaire on a computer. For each item in the 25-item concept list, the signer would first be presented with the Swedish translation of the word and then demonstrate their preferred sign for the meaning specified before the questionnaire administrator would continue to a subsequent page showing the animated `.gif`s. At this stage, the signer could choose to mark several signs as possible variants they would use. We collected data from 26 signers (12 female, 14 male; mean age 47; median age 49).[4] Their responses were compiled and sorted by signer metadata into a pilot study for evaluating the usefulness of the questionnaire. Data compilation was done with the statistical language R (R Core Team, 2015) and the data were plotted with the `ggplot2` package (Wickham, 2009).
In Figures 8–12 below, we illustrate the responses from our 26 Deaf primary SSL users for the meanings TIO ('ten'), NITTIO ('ninety'), and TORSDAG ('Thursday') across age groups (bins show decade of birth).

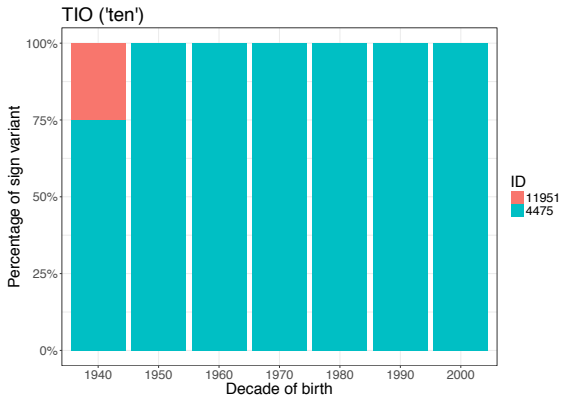


Figure 8: The relative distribution of the signs TIO (4475) and TIO(Y) (11951) ('ten')

Figure 8 confirms the pattern found in Figures 1–2, namely that TIO(Y) is less common overall, and that it is mainly confined to the usage of older signers (only present in the oldest age group here).
From Figure 9, we see that the sign form NITTIO(E) is becoming more popular over time, which we again would predict from knowing that NITTIO(4) is the older sign variant, and also a suppletive numeral by not being based on the sign NIO ('nine') as other tens.
Lastly, Figure 12 gives us the clearest example of a form change over time. The sign TORSDAG(L) is mostly used by signers born before 1980, and TORSDAG(Lb) mostly by

[3]We have converted all sign videos (`.mp4`) in the SSLD to `.gif` format with the intention to make these available in the online database for the public in a future release. The purpose of adding `.gif` files is to allow for sharing signs in, e.g., social media more easily. However, the original `.mp4` videos will remain the primary format in the sign videos in the SSLD.

[4]In total, 32 people responded to our questionnaire. Here, we only report the results of the respondents identifying as Deaf.

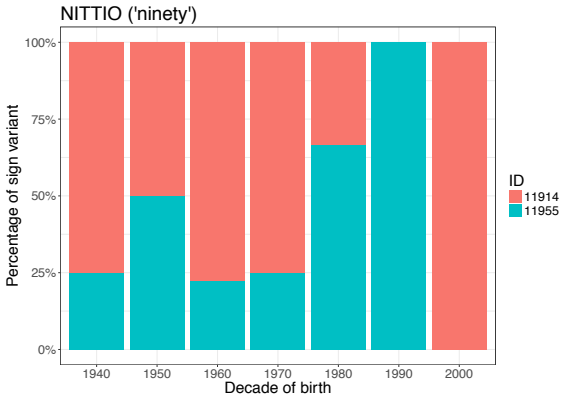


Figure 9: The relative distribution of the signs NITTIO(4) (11914) and NITTIO(E) (11955) ('ninety')

Figure 10: The sign NITTIO(E) (11955) ('ninety')

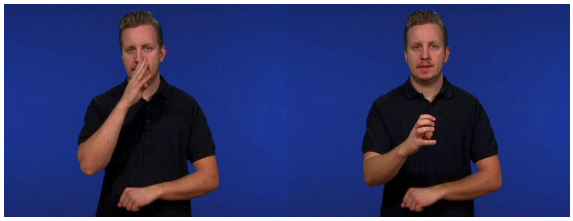

Figure 11: The sign NITTIO(4) (11914) ('ninety')

signers born after 1970, with the 1970s and 1980s as the transitional decades.
Here, we have shown three examples of meanings for which there was an suspected diachronic change in which sign variants are used, and the hypotheses about their distribution and change in distribution are tentatively confirmed by our data. Although these three examples give us an indication of patterns, we would need to gather much more data in order for it to be reliable and thus useful. We have the intention of continuing with the questionnaire type data collection, but this time entirely online in order to reach a larger set of community members more efficiently.

## 5. Conclusions

In this paper, we have described two of the main methods that we use in order to crowdsource data about SSL directly from the SSL community.
In the case of our Facebook group, we are able to collect

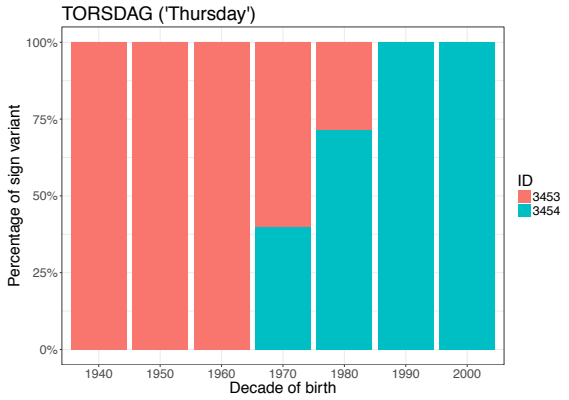


Figure 12: The relative distribution of the signs TORSDAG(L) (3453) and TORSDAG(Lb) (3454) ('Thursday')

Figure 13: The sign TORSDAG(L) (3453) ('Thursday')

Figure 14: The sign TORSDAG(Lb) (3454) ('Thursday')

signs directly, disseminate information about (new) signs to various subgroups in the community (Deaf, hard-of-hearing, hearing) – both ourselves and in interaction with other community members – and also to collect judgments about signs already in the SSLD or signs that could be included. The on-going documentation process enables us to quickly add new signs in consultation with the community members when needed, and also update or edit entries accordingly.

With our direct interaction with the community members at Deaf events (e.g., the Deaf Day), we are able to collect data with more extensive metadata about each signer, target specific individuals or groups that we need for our documentation work, and ask more qualitative questions about signs and sign usage. Thus, it complements the more passive (and massive) interaction taking place in our Facebook group. The results from our pilot study questionnaire shows that it could be a useful method for moving on to a larger online variation study, using a similar type of questionnaire with respondents being able to provide data from their own homes rather than requiring a direct interaction. This would further utilize the benefits of crowdsourcing, which can lead to a lot of new data in a very short time.

Crowdsourcing for the SSLD is a useful and rapid method for enhancing our language resource. It is especially convenient using the online community (e.g., our Facebook group) since it easily targets a large group of community members simultaneously. However, data collected through offline methods give may provide a more qualitative approach on sign variants, interacting with individuals directly, and allows for better control over signer metadata and responses, by being manually annotated – aside from giving the dictionary team an opportunity to personally meet and interact with the deaf community, which is an important aspect of any language documentation work.

## 6. Acknowledgments

We wish to thank the members of the SSL community who, through channels like our Facebook group and the Deaf Day, help us enrich and enhance our language resources part of our documentation and research of Swedish Sign Language. We are also grateful to the reviewers for helpful comments on a previous version of this paper.

## 7. Bibliographical References

Börstell, C. and Östling, R. (2016). Visualizing lects in a sign language corpus: Mining lexical variation data in lects of Swedish Sign Language. In Eleni Efthimiou, et al., editors, *Proceedings of the 7th Workshop on the Representation and Processing of Sign Languages: Corpus Mining*, pages 13–18, Paris. European Language Resources Association (ELRA).

Hedberg, T., Lindström, M., Lyxell, T., Mumtaz, S. A., and Tersmeden-Lundquist, E. (1998). *Svenskt teckenspråkslexikon*. Sveriges dövas riksförbund (SDR), Leksand.

Mesch, J. and Wallin, L. (2012). From meaning to signs and back: Lexicography and the Swedish Sign Language Corpus. In Onno Crasborn, et al., editors, *Proceedings of the 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon [Language Resources and Evaluation Conference (LREC)]*, pages 123–126, Paris. European Language Resources Association (ELRA).

Mesch, J., Wallin, L., and Björkstrand, T. (2012). Sign Language Resources in Sweden: Dictionary and Corpus. In Onno Crasborn, et al., editors, *Proceedings of the 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon [Language Resources and Evaluation Conference (LREC)]*, pages 127–130, Paris. European Language Resources Association (ELRA).

R Core Team, (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Stamp, R., Schembri, A., Fenlon, J., Rentelis, R., Woll, B., and Cormier, K. (2014). Lexical variation and change in British sign language. *PLoS ONE*, 9(4).

Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

## 8. Language Resource References

Börstell, C. and Östling, R. (2018). SSL-lects: An online interface for plotting lexical variation data from the Swedish Sign Language Corpus. Department of Linguistics, Stockholm University. mumin.ling.su.se/cgi-bin/ssllects.py.

Mesch, J., Wallin, L., Nilsson, A.-L., and Bergman, B. (2012). Dataset. Swedish Sign Language Corpus project 2009–2011 (version 1).

Mesch, J. (2018). Annotated files for the Swedish Sign Language Corpus (Version 8). Department of Linguistics, Stockholm University. URL.

Svenskt teckenspråkslexikon. (2018). Swedish Sign Language Dictionary online. Department of Linguistics, Stockholm University. teckenspraklexikon.su.se.